SUSMAN GODFREY L.L.P.

September 6, 2024

Hon. Ona T. Wang
Southern District of New York
500 Pearl Street
New York, New York 10007

> Re: *The New York Times Company v. Microsoft Corporation, et al.*,
> Case No.: 23-cv-11195-SHS: Opposition to OpenAI's Motion to Compel

Dear Magistrate Judge Wang:

Plaintiff The New York Times Company ("The Times") opposes OpenAI's September 3 request to compel the production of documents. Dkt. 236. As an initial matter, OpenAI misstates what The Times agreed to produce: The Times has already agreed to produce the materials related to its use of, and positions about, generative AI that are relevant to this case. OpenAI's motion seeks broad discovery into documents about its competitors' products, "copyright," and "AI" generally that are simply irrelevant to the issues before this Court.

1. The Times Already Agreed to Produce a Substantial Number of Documents about Generative AI.

The Times did not "refuse to produce any discovery relating to any tools other than those created by OpenAI," as OpenAI claims. Dkt. 236 at 2. Notwithstanding concerns about overbreadth and lack of relevance, in order to avoid disputes, The Times agreed to produce documents regarding The Times's A.I. Initiatives program (OAI RFP 47), documents regarding The Times's use of both OpenAI and Microsoft's generative AI tools in reporting or presentation of content (OAI RFPs 49, 83), documents sufficient to show any Times trainings about generative AI (OAI RFP 46), and documents regarding Times policies about OpenAI and Microsoft's generative AI tools (OAI RFP 44, MS RFPs 29-30). *See* Ex. D; Ex. 1 (summary exhibit of relevant discovery responses); Ex. 5.

2. The Times's Use of Third Parties' Products is Not Relevant to Any Fair Use Defense.

OpenAI argues that it needs discovery into The Times's use of "generative AI generally"— as in, generative AI tools *not* created by OpenAI or Microsoft—to show the "public benefit" of its copying of Times content. But this argument misunderstands how the "public benefit" analysis works. It is not a stand-alone question, but instead is part of the fourth fair use factor: "the effect of *the use* upon the potential market for or value of the copyrighted work." 17 U.S.C. § 107(4) (emphasis added). Specifically, the court may assess "the public benefits *the copying* will likely produce." *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 35 (2021) (emphasis added). Here, "the use" and "the copying" at issue is *OpenAI and Microsoft's* unauthorized use and copying of The Times's asserted works, and "the potential market" is the market for the same Times works.

Put differently, it is Microsoft and OpenAI's burden to prove that, by using Times works without permission, they have provided a benefit to the public. *See Hachette Book Group, Inc. v. Internet Archive*, -- F.4th --, 2024 WL 4031751, at *14 (2d Cir. Sep. 4, 2024) (holding that the

September 6, 2024
Page 2

defendant bears the burden of proving the "effect on the market" fair use factor). The use or non-use of third-party generative AI tools by The Times is wholly irrelevant to that inquiry. To the extent a look at broader industry conduct is relevant, it is as part of an analysis about the impact Defendants' copying has on publishers' businesses and whether it decreases their incentive to create works in the first place. *See Sega Ent. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1523 (9th Cir. 1992) (internal citations omitted) (stating that a fair use analysis may consider whether, if "the challenged use should become widespread, it would adversely affect the potential market for the copyrighted work,' by diminishing potential sales, interfering with marketability, or usurping the market").

There is a very good reason why courts do not analyze whether copying by non-parties provides a "public benefit." If allowed, it would turn a case against particular defendants into a series of mini-trials about the conduct of *other* companies and the nature of their AI products. OpenAI's argument, at its core, is that this discovery will help it prove that its copying has a "public benefit" because other companies' products provided a benefit to The Times (presumably beyond any money The Times has to pay to use them). This would require OpenAI to prove (among other things) that those third-party tools *also* infringed the same Times works, the character of their unauthorized use was the same as OpenAI's, and they commercialized their products in the same way. And even if OpenAI's trials-within-a-trial were successful, The Times's use of any such third-party tools would still not be relevant to evaluating the effects of OpenAI's infringing tools on the market for Times works or on the public interest. The Times does not act as a consumer in the market for its own copyrighted works.

3. <u>The Times's Own Tools are Irrelevant.</u>

Discovery into The Times's creation of AI tools is also irrelevant. Even so, The Times already agreed to produce documents and communications relating to its A.I. Initiatives program—the same program OpenAI cites as a basis for seeking discovery on this topic. *See* Dkt. 236 at 3 n.5.

Yet that is not enough for OpenAI. RFP 51 seeks documents about the training of a generative AI tool that The Times introduced for its advertising business, a use that is clearly not at issue in this case. And the other discovery OpenAI demands is not about generative AI at all. RFP 50 seeks documents about an AI tool that was designed to examine satellite images to trace the path of the "Chinese spy balloon" spotted in 2023 for reporting purposes.[1] The AI satellite tool uses AI, not generative AI, which even OpenAI does not claim to have "pioneered and introduced to the world." Dkt. 236 at 2.

These requests do not have anything to do with whether OpenAI's unauthorized use of The Times's copyrighted content to train a generative chatbot that competes with The Times's journalism constitutes "fair use," nor do they have anything to do with The Times's damages

---

[1] *See* Eduardo Suarez, *New York Times publisher A.G. Sulzberger: "Our industry needs to think bigger"*, Reuters Institute (Feb. 19, 2024), https://reutersinstitute.politics.ox.ac.uk/news/new-york-times-publisher-g-sulzberger-our-industry-needs-think-bigger.

September 6, 2024
Page 3

(which OpenAI mentions but does not explain). And neither are actually directed at the information OpenAI claims to need. *See* Dkt. 236-3 (OpenAI arguing it is entitled to know whether The Times "has leveraged unlicensed data to create these tools," "admitted internally that there is no 'workable market' for licensing the data needed to create an effective language model," or "acknowledged that this technology could enhance the 'traditional business models that supported quality journalism'"). OpenAI does not explain how discovery into either tool could plausibly uncover relevant admissions about the feasibility of licensing its own works to build an LLM or the benefits of OpenAI's infringing use of Times works to journalism, or why any such information would not be discoverable in the numerous documents The Times agreed to produce.

4. OpenAI's Remaining Requests Are Overbroad.

Here again, The Times has already agreed to produce non-privileged communications with third parties about the relevant issues in this case: Defendants' use of Times content in their generative AI products, this litigation, and whether to license works to OpenAI. The Times also agreed to produce documents regarding The Times's attempts to license its works to OpenAI and Microsoft, Defendants' use of Times works to train, and OpenAI competing with Times works and creating substitutive products (OAI RFPs 1, 3, 16-17, 25, 40, 58, 87-89, 90, 92; MS RFPs 33, 50).[2] *See* Exs. 1-3, 5; Ex. D. OpenAI's demand for all communications regarding "copyright" and "AI" is overbroad and unsupported by any caselaw.[3]

As for RFPs 4-6 and 16, OpenAI claims that discovery into whether The Times knew third-party generative AI companies infringed its works and did nothing is relevant to show The Times recognizes that "such training is protected by fair use." Dkt. 236-3. Although this relevance hook is attenuated at best, The Times already agreed to produce documents concerning The Times's efforts to monitor for, investigate, or enforce infringement of its intellectual property rights (MS RFP 27, OAI RFP 65). Exs. 1-2, 5. The information OpenAI seeks is covered by these requests, and further discovery into The Times's knowledge of any other "alleged" use of Times works by third parties is unnecessary and overbroad. *See* Ex. D (RFPs 4-6, 16).

Respectfully submitted,

/s/ Ian B. Crosby
Ian B. Crosby
Susman Godfrey L.L.P.

/s/ Steven Lieberman
Steven Lieberman
Rothwell, Figg, Ernst & Manbeck

---

[2] *See* Ex. 3 (The Times's July 25, 2024 discovery correspondence with counsel for Microsoft agreeing to produce documents responsive to Microsoft RFPs 33 and 50).

[3] During meet and confer efforts on OpenAI's RFP 24, The Times offered to produce communications with those third parties relating to copyright infringement by OpenAI. Ex. 4. OpenAI declined.

September 6, 2024
Page 4

cc: All Counsel of Record (via ECF)